CLAIMS

What is claimed is:

1   1.   A method for managing memory used for query execution, the method comprising
2   the computer-implemented steps of:

3   allocating a buffer pool in a cache memory;

4   wherein the cache memory is managed by a cache memory manager;

5   wherein the buffer pool is managed by a buffer pool manager that is distinct from said

6   cache memory manager;

7   wherein the buffer pool includes a plurality of buffers;

8   in response to a need to allocate space in said cache memory for a query working set

9   of a query, the buffer pool manager allocating a buffer from the buffer pool to

10   the query working set.

1   2.   The method of Claim 1 wherein the step of allocating a buffer includes allocating a
2   buffer to store a frame buffer and a bind buffer associated with said query.

1   3.   The method of Claim 1 wherein the step of allocating a buffer pool includes
2   allocating a buffer pool in an L1 cache that resides on a processor.

1   4.   A method as recited in Claim 1, further comprising the step of:

2   prior to allocating said buffer pool, determining a size for said buffer pull by

3   calculating a total number of buffers to include in the buffer pool.

1   5.   A method as recited in Claim 4, wherein the step of calculating a total number of
2   buffers to include in the buffer pool further comprises the steps of:

3   calculating working set memory requirements of an average query;

4   sizing buffers in the buffer pool to accommodate the working set memory

5   requirements of the average query;

6   generating an estimate of how many queries will be executing at a given time; and

7   setting the number of buffers in the buffer pool to a number sufficient to

8   accommodate the estimated number of queries executed at a given time.

1 6.   A method as recited in Claim 4, wherein the step of calculating total number of

2 buffers to include in the buffer pool further comprises the steps of:

3   calculating working set memory requirements of queries executing during hot

4   sections;

5   calculating working set memory requirements of frequently used queries;

6   sizing buffers in the buffer pool to accommodate the working set memory

7   requirements of the queries executing during hot sections and the working set

8   memory requirements of the frequently used queries; and

9   setting the number of buffers in the buffer pool to a number sufficient to

10   accommodate queries executed during a hot section.

1 7.   A method as recited in Claim 1, wherein the buffer pool manager does not allocate a

2 buffer from the buffer pool to a query working set if the working set is larger than the size of

3 buffers within said buffer pool.

1 8.   A method as recited in Claim 1, further comprising the step of:

2   generating a buffer map table;

3   wherein the map table points to free and used buffers in the buffer pool; and

4   wherein the buffer pool manager checks the map table for free buffers in the buffer

5   pool prior to allocating buffers in the buffer pool.

1   9.   A method as recited in Claim 8, wherein the buffer pool manager allocates a buffer

2   from the buffer pool to a query working set if the map table shows that a buffer is free in the

3   buffer pool.

1   10.   A method as recited in Claim 8, wherein the buffer pool manager does not allocate a

2   buffer from the buffer pool to a query working set if the map table shows that no buffers are

3   free in the buffer pool.

1   11.   A method as recited in Claim 8, wherein the buffer pool manager frees a buffer in the

2   buffer pool, without removing the buffer from the buffer pool, after a query releases its

3   allocated buffer by marking a location in the map table that corresponds to the released

4   buffer as free.

1   12.   The method of Claim 1 wherein said buffer pool is a first buffer pool, and the method

2   further comprises the steps of:

3   allocating in said cache memory a plurality of buffer pools;

4   wherein said plurality of buffer pools include said first buffer pool and at least one

5   other buffer pool;

6   wherein each buffer pool of said plurality of buffer pools is comprised of a plurality

7   of buffers;

8   wherein each buffer pool of said plurality of buffer pools has a characteristic that

9   differs from each other buffer pool of said plurality of buffer pools; and

10   said buffer pool manager determines which buffer pool of said plurality of buffer

11   pools should be used to store said query working set based on the

12   characteristics of said buffer pools.

1   13.   The method of Claim 12 wherein buffers in each buffer pool of the plurality of buffer

2   pools have a different size than buffers in the other buffer pools in said plurality of buffer

3   pools.

1   14.   A method as recited in Claim 12, further comprising, prior to allocating said plurality

2   of buffer pulls, calculating total number of buffers to include in each buffer pool.

1   15.   A method as recited in Claim 14, wherein the step of calculating a total number of

2   buffers to include in each buffer pool further comprises the steps of:

3   calculating working set memory requirements of queries executing during hot

4   sections;

5   grouping the queries in the hot sections into at least two groups based on common

6   working set memory requirements;

7   sizing buffers in a particular buffer pool to accommodate the working set memory

8   requirements of queries in one group; and

9   setting the number of buffers in the particular buffer pool to a number sufficient to

10   accommodate the number of queries executed during a hot section in the one

11   group.

1   16.   A method as recited in Claim 15, wherein the number of buffers in a buffer pool is set

2   so that some queries may be allocated memory from another buffer pool that has buffers that

3   are larger than needed and are unused during a particular hot section.

1   17.   A method as recited in Claim 12, wherein the buffer pool manager does not allocate a

2   buffer from a buffer pool to a query working set if the query working set is larger than the

3   largest buffer's size.

1   18.   A method as recited in Claim 12, further comprising the step of:

2   providing a buffer map table;

3   wherein the map table points to free and used buffers in the buffer pools; and

4   wherein the buffer pool manager checks the map table for free buffers in a buffer pool

5   prior to allocating buffers in the buffer pool.

1   19.   A method as recited in Claim 18, wherein the buffer pool manager allocates a buffer

2   from the buffer pool to a query working set if the map table shows that a buffer is free in the

3   buffer pool.

1   20.   A method as recited in Claim 18, wherein the buffer pool manager does not allocate a

2   buffer from the buffer pool to a query working set if the map table shows that no buffers are

3   free in the buffer pool.

1   21.   A method as recited in Claim 18, wherein the buffer pool manager frees a buffer in

2   the buffer pool after a query releases its allocated buffer by marking a location in the map

3   table that corresponds to the released buffer as free.

1   22.   A computer-readable medium carrying one or more sequences of instructions which,

2   when executed by one or more processors, causes the one or more processors to perform the

3   method recited in Claim 1.

1   23.   A computer-readable medium carrying one or more sequences of instructions which,

2   when executed by one or more processors, causes the one or more processors to perform the

3   method recited in Claim 2.

1   24.   A computer-readable medium carrying one or more sequences of instructions which,
2   when executed by one or more processors, causes the one or more processors to perform the
3   method recited in Claim 3.

1   25.   A computer-readable medium carrying one or more sequences of instructions which,
2   when executed by one or more processors, causes the one or more processors to perform the
3   method recited in Claim 4.

1   26.   A computer-readable medium carrying one or more sequences of instructions which,
2   when executed by one or more processors, causes the one or more processors to perform the
3   method recited in Claim 5.

1   27.   A computer-readable medium carrying one or more sequences of instructions which,
2   when executed by one or more processors, causes the one or more processors to perform the
3   method recited in Claim 6.

1   28.   A computer-readable medium carrying one or more sequences of instructions which,
2   when executed by one or more processors, causes the one or more processors to perform the
3   method recited in Claim 7.

1   29.   A computer-readable medium carrying one or more sequences of instructions which,
2   when executed by one or more processors, causes the one or more processors to perform the
3   method recited in Claim 8.

1   30.   A computer-readable medium carrying one or more sequences of instructions which,
2   when executed by one or more processors, causes the one or more processors to perform the
3   method recited in Claim 9.

1   31.   A computer-readable medium carrying one or more sequences of instructions which,

2   when executed by one or more processors, causes the one or more processors to perform the

3   method recited in Claim 10.

1   32.   A computer-readable medium carrying one or more sequences of instructions which,

2   when executed by one or more processors, causes the one or more processors to perform the

3   method recited in Claim 11.

1   33.   A computer-readable medium carrying one or more sequences of instructions which,

2   when executed by one or more processors, causes the one or more processors to perform the

3   method recited in Claim 12.

1   34.   A computer-readable medium carrying one or more sequences of instructions which,

2   when executed by one or more processors, causes the one or more processors to perform the

3   method recited in Claim 13.

1   35.   A computer-readable medium carrying one or more sequences of instructions which,

2   when executed by one or more processors, causes the one or more processors to perform the

3   method recited in Claim 14.

1   36.   A computer-readable medium carrying one or more sequences of instructions which,

2   when executed by one or more processors, causes the one or more processors to perform the

3   method recited in Claim 15.

1   37.   A computer-readable medium carrying one or more sequences of instructions which,

2   when executed by one or more processors, causes the one or more processors to perform the

3   method recited in Claim 16.

1   38.   A computer-readable medium carrying one or more sequences of instructions which,

2   when executed by one or more processors, causes the one or more processors to perform the

3   method recited in Claim 17.

1   39.   A computer-readable medium carrying one or more sequences of instructions which,

2   when executed by one or more processors, causes the one or more processors to perform the

3   method recited in Claim 18.

1   40.   A computer-readable medium carrying one or more sequences of instructions which,

2   when executed by one or more processors, causes the one or more processors to perform the

3   method recited in Claim 19.

1   41.   A computer-readable medium carrying one or more sequences of instructions which,

2   when executed by one or more processors, causes the one or more processors to perform the

3   method recited in Claim 20.

1   42.   A computer-readable medium carrying one or more sequences of instructions which,

2   when executed by one or more processors, causes the one or more processors to perform the

3   method recited in Claim 21.